IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **41** (2008) 415001 (13pp) doi:10.1088/1751-8113/41/41/415001

Broad edge of chaos in strongly heterogeneous Boolean networks

Deok-Sun Lee[1] **and Heiko Rieger**[2]

[1] Center for Complex Network Research and Department of Physics, Northeastern University, Boston, MA 02115, USA
[2] Theoretische Physik, Universität des Saarlandes, 66041 Saarbrücken, Germany

Received 14 March 2008, in final form 25 July 2008
Published 19 September 2008
Online at stacks.iop.org/JPhysA/41/415001

Abstract

The dynamic stability of the Boolean networks representing a model for the gene transcriptional regulation (Kauffman model) is studied by calculating analytically and numerically the Hamming distance between two evolving configurations. This turns out to behave in a universal way close to the phase boundary only for in-degree distributions with a finite second moment. In-degree distributions of the form $P_d(k) \sim k^{-\gamma}$ with $2 < \gamma < 3$, thus having a diverging second moment, lead to a slower increase of the Hamming distance when moving towards the unstable phase and to a broadening of the phase boundary for finite N with decreasing γ. We conclude that the heterogeneous regulatory network connectivity facilitates the balancing between robustness and evolvability in living organisms.

PACS numbers: 89.75.Hc, 64.60.Cn, 05.65.+b, 02.50.−r

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Complete genome sequencing and the analysis of the binding of transcriptional regulators to specific promoter sequences have uncovered the global organization of the gene transcriptional regulatory network in well-studied organisms such as *Escherichia coli* [1] and yeast *Saccharomyces cerevisiae* [2]. The gene network describes a directed relationship—regulation—between different genes and its architecture is characterized by broad connectivity distributions [1–4], over-representation of selected motifs [5] and so on. These features are rarely found in random networks, probably being the consequence of evolutionary selection. Therefore, illuminating the functional characteristics associated with those discovered structural features can help trace their origins. In this work, we show that heterogeneous connectivity can facilitate the balancing between dynamical stability and instability. Both robustness and evolvability are essential for living organisms, which achieve their specific

phenotype by their gene expression program [6]. Thus the transcriptional regulatory network should be organized in a way that supports the coexistence of these apparently contradictory properties and from this perspective, it has been proposed that the gene network should be at the boundary between stable and unstable phases, called the edge of chaos [7]. The question then arises: what are the characteristics of the network architecture that can support the requirement to be located at the edge of chaos? A simple model incorporating recently available information turns out to be useful to answer this question.

The Kauffman model [7] was used in the past to study the gene network dynamics which is far from completely known because of its complexity. In this model, each node has a Boolean variable, 1 or 0, the discretized expression level evolving, regulated by other K nodes according to the quenched rules that are randomly distributed with a parameter p. In spite of these simplifications involved, the model revealed detailed relations between the dynamical stability against perturbations and the network architecture [7, 8]. Moreover, distinct attractors in the configuration space are considered as corresponding to different cell types in a given organism and thus its scaling with the number of genes (nodes) across different organisms has been of great interest [7, 9–14]. Empirically, the number of cell-type scales as the square-root of the number of genes and the same scaling relation was believed to hold between the number of attractors and the number of nodes in the Kauffman model at the critical point $p_c(K)$, supporting the hypothesis that living organisms should be between order and chaos [7]. Recently, however, it was found that under-sampling effects may hamper numerical enumeration of distinct attractors [10, 11] and further investigations demonstrated that the total number of attractors grows faster than any power law with the system size [12, 13]. On the other hand, it was also reported that attractors stable against deviation from synchronous update show sub-linear scaling behavior [14].

Recent investigations of real gene networks suggest generalization of the original Kauffman model. First, the distribution of the regulating rules is structured showing a bias towards the canalyzing functions [15, 16]. Second, the number of links or degree is not constant but different from node to node, resulting in broad degree distributions [17]. In the gene regulatory networks of *E. coli* [3, 5, 18] and yeast [2, 4, 19–21], the distributions of out-degree (number of target genes for each regulator) and in-degree (number of regulators for each target gene) were not delta-functions but shown to take power-law or exponential-decaying form, respectively, although true asymptotic behavior was hard to discern due to finite-size effects. While the effects of the structured distribution of regulating rules have been intensively studied [16, 19, 22], it remains to show how the heterogeneous connectivity affects the dynamical stability [23, 24].

We consider the Kauffman model on directed networks with general in- and out-degree distributions and compare two evolving dynamical configurations by computing their Hamming distance, to determine whether a given network is dynamically stable (zero distance) or unstable (nonzero distance) against perturbations. The critical point of the Boolean networks with power-law degree distributions has recently been studied [24]. In the present work, we show quantitatively how the Hamming distance behaves near the critical point, which will provide a deeper understanding of the critical phenomena of Boolean networks with heterogeneous connectivity patterns and insights into the interplay of structure and dynamics in living organisms. The Hamming distance for infinite system size (thermodynamic limit) can be computed by the method presented in [18] and we here present a detailed description of the method along with a discussion on the effects of correlation between in- and out-degree of the same node. Then, more importantly, we extend the method to derive the Hamming distance for finite system size, which enables us to check the analytic predictions with numerical simulation results. Our main result is that for in-degree distributions with a diverging second

moment the Hamming distance increases very slowly when moving from the phase boundary towards the unstable phase and the width of the boundary in finite-size systems is very broad. This indicates that strongly heterogeneous genetic networks have a large capacity to stay at the edge of chaos when their structural and functional organization is subject to variation.

The paper is organized as follows. We introduce the Kauffman model for Boolean networks in section 2. In section 3, the annealed approximation is described and used to compute the Hamming distance, which reveals different phases of the Boolean networks. The finite-size effects on the critical phenomena of the Boolean networks are derived using the annealed approximation in section 4. Finally, the results are summarized and discussed in section 5.

2. Model and Hamming distance

In the Kauffman model, the dynamical configuration of N Boolean variables at time t, $\Sigma(t) = \{\sigma_i(t)|i = 1, 2, \ldots, N\}$, is updated in parallel as

$$\sigma_i(t+1) = f_i(\Sigma_i(t)), \tag{1}$$

where σ_i for each i takes 1 or 0 and $\Sigma_i(t) = \left\{\sigma_{i_1}(t), \sigma_{i_2}(t), \ldots, \sigma_{i_{k_i}}(t)\right\}$ denotes the configuration at time t of the k_i regulators $\mathcal{R}_i = \left\{i_1, i_2, \ldots, i_{k_i}\right\}$, of the node i. The functional dependences between nodes via $\{f_i(\Sigma_i)|i = 1, 2, \ldots, N\}$ constitute a directed network in which two nodes i and j are connected with a directed edge (i, j) if $j \in \mathcal{R}_i$, where (i, j) is an outgoing edge of node j and an incoming edge of node i. The quenched, i.e., time-independent, regulating rules are random Boolean functions, i.e., they are chosen randomly such that $f_i(\Sigma_i)$ for a given Σ_i is 1 with probability $0 \leqslant p \leqslant 1$ and 0 with probability $1 - p$. The parameter p deviating from $1/2$ indicates an asymmetry between expressed (1) and nonexpressed (0) states of a gene.

We focus on the following question: if one starts at time $t = 0$ with two randomly chosen configurations, Σ and $\hat{\Sigma}$ with $\hat{\sigma_j} \neq \sigma_j$ for all j, that is, all node states perturbed (altered), how many nodes remain perturbed at time $t > 0$? The fraction of these perturbed nodes or the Hamming distance between Σ and $\hat{\Sigma}$ at time t is defined as

$$H(t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\sigma_i(t), 1-\hat{\sigma_i}(t)} \tag{2}$$

with $\delta_{a,b}$ being 1 for $a = b$ and 0 otherwise. The Hamming distance may vary between 0 and 1 depending on the dynamic asymmetry parameter p. We will see in the following section that the value of the Hamming distance in the stationary state may display a transition from zero to a nonzero value as the network architecture and the parameter p are varied.

3. Annealed approximation and phase transition of Boolean networks

In this section, we investigate the phase diagram of the Kauffman Boolean network defined in the previous section by computing analytically and numerically the Hamming distance for infinite system size. This allows us to understand different phases of the Boolean networks determined by network structure and the parameter of dynamic asymmetry. Some of the results presented in this section are also found in [18].

3.1. Annealed approximation

A recursion relation for the Hamming distance between consecutive time steps is obtained by the 'annealed' approximation [8]. While the regulation rule f_i and the regulators $\mathcal{R}_i$ are fixed for each node i in the original model, one assigns them randomly to every node at every time step, keeping the in-degrees and the out-degrees, in the annealed approximation. Then the evolution of the Hamming distance $H_{k,q}(t)$ for the nodes with in-degree k and out-degree q is given by $H_{k,q}(t+1) = \lambda\big[1 - \big(1 - \sum_{k',q'} q' P_d(k', q') H_{k',q'}(t)/\langle q\rangle\big)^k\big]$, where $\lambda \equiv 2p(1-p)$, $P_d(k', q')$ is the joint distribution of k' and q', and $\langle q\rangle = \sum_{k,q} q P_d(k, q)$. The correlation between the degrees of neighboring nodes, $\{k, q\}$ and $\{k', q'\}$ is ignored in this formalism but will be discussed in section 5. The parameter λ ranging from 0 to 1/2 is the probability that f_i yields different outputs for Σ_i and $\hat{\Sigma}_i$ given differently and the term within the brackets represents the probability of the latter. Note that the degree distribution for the regulators is weighted by their out-degrees. If we introduce $\bar{H}(t) \equiv \sum_{k,q}[q P_d(k, q)/\langle q\rangle] H_{k,q}(t)$, it is obtained self-consistently and in turn $H(t) = \sum_{k,q} P_d(k, q) H_{k,q}(t)$ is computed as follows [18]:

$$\begin{aligned} \bar{H}(t+1) &= \lambda \sum_{k,q} \frac{q P_d(k,q)}{\langle q\rangle}[1 - (1 - \bar{H}(t))^k], \\ H(t+1) &= \lambda \sum_{k} P_d(k)[1 - (1 - \bar{H}(t))^k], \end{aligned} \tag{3}$$

where $P_d(k) = \sum_q P_d(k, q)$ is the in-degree distribution. In the original Kauffman model where the in-degree is fixed to $k = K$, equation (3) reduces to $H(t+1) = \lambda[1 - (1 - H(t))^K]$ [8].

3.2. Ordered and chaotic phases

The limiting value $H = \lim_{t\to\infty} H(t)$ can characterize the system's response to dynamical perturbations. Replacing $H(t)$ and $\bar{H}(t)$ with H and $\bar{H}$, respectively, and expanding the first line in equation (3) for small $\bar{H}$, one finds that $H = 0$ and $\bar{H} = 0$ for $\lambda < \lambda_c$ and $H > 0$ and $\bar{H} > 0$ for $\lambda > \lambda_c$, where the critical point λ_c depends on the network topology via

$$\lambda_c \equiv K^{-1} \qquad \text{with} \quad K \equiv \sum_{k,q} \frac{kq P_d(k,q)}{\langle q\rangle}. \tag{4}$$

That is, the system is in the ordered phase, any perturbation making no effect on the system eventually, when $\lambda < \lambda_c$. On the other hand, the system does not remain in the same stationary state but shifts to another stationary state triggered by a perturbation when $\lambda > \lambda_c$.

The quantity K may be considered as the average in-degree of regulators weighted by their out-degrees. Introducing the conditional average of out-degree, $q_k \equiv \sum_q q P_d(k, q)/P_d(k)$, we can rewrite the quantity K as $K = \sum_k k(q_k/\langle q\rangle) P_d(k)$ and also the first relation in equation (3) as

$$\bar{H}(t+1) = \lambda \sum_{k} \frac{q_k}{\langle q\rangle} P_d(k)[1 - (1 - \bar{H}(t))^k]. \tag{5}$$

If q_k is independent of k or (more strongly) the in-degree and the out-degree of a node is not correlated statistically, it follows that $q_k = \langle q\rangle$ for all k, K reduces to the conventional average in-degree $\langle k\rangle \equiv \sum_k k P_d(k)$, and $H(t) = \bar{H}(t)$. The analyses of the transcriptional regulatory networks of *E. coli* and yeast show no significant variation of q_k with k [25] and so we will assume in the following that $q_k = \langle q\rangle$ for all k. In section 4.3, we will discuss

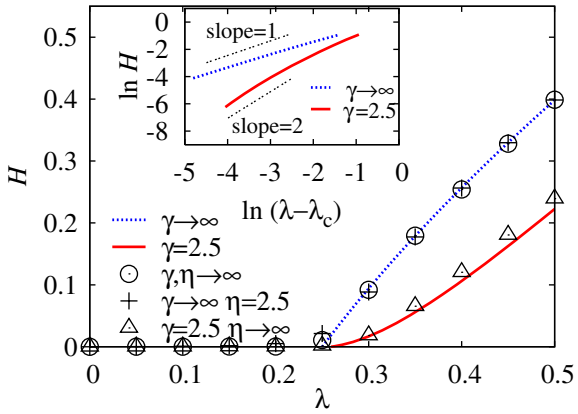

Figure 1. Hamming distance H for the Kauffman model as a function of λ. The points are the simulation results for model networks [26] with $N = 10^4$ and $\langle k\rangle = 4$ that have $P_d(k,q) = g_\gamma(k)g_\eta(q)$, where $g_a(x) \sim x^{-a}$ for a finite and $g_\infty(x) = \langle k\rangle^x \mathrm{e}^{-\langle k\rangle}/x!$. The Hamming distance H was first averaged over time between $t = 100$ and $t = 200$, and then averaged over 1000 different realizations of networks and regulating rules for which $H \neq 0$. The lines are the numerical solutions to equation (3) with $H(t) = \tilde{H}(t) = H$ for all t and the same degree distributions as in the simulation used. (Inset) Plots of $\ln H$ versus $\ln(\lambda - \lambda_c)$ with $\lambda_c = 0.25$. The slopes agree with equation (8).

how our results for the critical phenomena would be changed by the k-dependence of q_k. Under this assumption ($q_k = \langle q\rangle$), the Hamming distance $H(t)$ depends only on the in-degree distribution $P_d(k)$ and the dynamics parameter λ.

It has been shown that the scaling behavior of the average number of attractors with the system size remains to be the same for different out-degree distributions such as uniform, exponential and power-law one [23]. Our analysis based on the annealed approximation suggests further the irrelevance of the out-degree distribution to the Hamming distance. To confirm this as well as check the validity of the annealed approximation or equation (3), we performed simulations of the Kauffman model defined in section 2 on an ensemble of model networks constructed as follows [26]: (i) each of N nodes has two indices i_{in} and i_{out}, which run from 1 to N respectively. The two indices are independent. (ii) Choose a node A with index i_{out} with probability $i_{\mathrm{out}}^{-\alpha_{\mathrm{out}}} \big/ \sum_j j^{\alpha_{\mathrm{out}}}$. (iii) Choose a node B indexed i_{in} with probability $i_{\mathrm{in}}^{-\alpha_{\mathrm{in}}} \big/ \sum_j j^{\alpha_{\mathrm{in}}}$. (iv) Assign a link from the node A to B unless they are connected. (v) Repeat (ii) and (iii) until the total number of links is L. The generated networks have N nodes, L links, and degree distribution given by $P_d(k,q) = P_d(k)P_d(q)$, where the in-degree distribution takes the form $P_d(k) \sim k^{-\gamma}$ and the out-degree distribution takes the form $P_d(q) \sim q^{-\eta}$ with $\gamma = 1 + 1/\alpha_{\mathrm{in}}$ and $\eta = 1 + 1/\alpha_{\mathrm{out}}$. It is then obvious that $q_k = \langle q\rangle$ for all k. When $\alpha_{\mathrm{in}} = 0$ and thus all the nodes can have an incoming link with equal probability, the in-degree distribution becomes a Poisson one, $P_d(k) = \langle k\rangle^k e^{-\langle k\rangle}/k!$. Thus the degree distribution may take power-law or Poissonian form depending on the values of α_{in} and α_{out} corresponding to scale-free (SF) networks or completely random networks, respectively. The simulation results (data points) shown in figure 1 are compared with the numerical solutions (lines) to equation (3), the annealed approximation, which show a good agreement and support the validity of the annealed approximation. Also it is shown that the Hamming distance is the same for different out-degree distributions.

The implication of equation (3) for SF networks has been discussed in [24], where $P_d(k) = k^{-\gamma}/\zeta(\gamma)$ for $k = 1, 2, \ldots$ with $\zeta(x)$ the Rieman–Zeta function. Based on the result that $\langle k\rangle = \zeta(\gamma-1)/\zeta(\gamma) < 2$ for $\gamma > 2.47875\ldots$, it was claimed [24] that the abundance of SF networks with $2 < \gamma < 2.5$ in nature and society can be attributed to the presence of both phases, stable and unstable, only in such networks. However, the values of $\langle k\rangle$ and γ do not show such strong correlation in real networks. For instance, the average degree $\langle k\rangle$ ranges from 2.57 (Internet router network) to 28.78 (movie actor network) although the degree exponent γ lies between 2 and 3 [17], which is possible due to the power-law behavior observed only asymptotically. We will show in the following section that the root for the dynamical advantage of SF network lies elsewhere.

3.3. Critical exponents

Here we address the behavior of the Hamming distance around λ_c for infinite system size. In that regime of λ, the Hamming distance is very small and its increase with $\lambda - \lambda_c$ can be characterized by a scaling exponent. This critical behavior of the Hamming distance is of interest to us because it shows how the network topology is related to the system's dynamic response.

When the moments $\langle k^n\rangle = \sum_k P_d(k)k^n$ for all $n > 0$ are finite, equation (3) can be written as

$$H \simeq \lambda \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n!}\langle k^n\rangle H^n. \tag{6}$$

Keeping the leading terms, we find that $H \simeq (\lambda/\lambda_c)H - \lambda\langle k^2\rangle H^2/2$, which gives

$$H \sim \Delta$$

with $\Delta \equiv \lambda/\lambda_c - 1$ for $0 < \Delta \ll 1$. This result can be represented as $H \sim \Delta^\beta$ with $\beta = 1$, where we introduced the critical exponent β.

On the other hand, if $P_d(k) \simeq ck^{-\gamma}$ for $k \gg 1$ with c a constant, $\langle k^n\rangle$ diverges as $ck_{\max}^{n+1-\gamma}/(n+1-\gamma)$ for $n \geqslant \lceil\gamma\rceil - 1$ with $k_{\max}$ the largest in-degree and $\lceil x\rceil$ denoting the smallest integer not smaller than x. Applying the relation $\sum_{k>k_{\max}} P_d(k) \sim 1/N$ from the extreme value statistics [27], one can see that $k_{\max}$ scales as $N^{1/(\gamma-1)}$. The diverging terms in equation (6) have alternating signs and lead to non-analytic terms in H as described below [28]. For small H, equation (3) reads as $H \simeq \lambda\sum_k P_d(k)[1 - \mathrm{e}^{-kH}]$ and recalling the power-law form of $P_d(k)$, $P_d(k) \simeq ck^{-\gamma}$, we can utilize the fact that the Mellin transform of a function $F(\gamma, H) \equiv \sum_{k=1}^{\infty} k^{-\gamma}\,\mathrm{e}^{-kH}$ is given by $\mathcal{F}(\gamma, s) = \int_0^\infty F(\gamma, H)H^{s-1}\,\mathrm{d}H = \Gamma(s)\zeta(s+\gamma)$ with $\Gamma(x)$ the Gamma function [28]. The inverse transform of $\mathcal{F}(\gamma, s)$ is then represented in terms of the poles of the Rieman–Zeta function and the Gamma function, which gives $F(\gamma, H) = \int_{c-\mathrm{i}\infty}^{c+\mathrm{i}\infty} \mathcal{F}(\gamma, s)H^{-s}\,\mathrm{d}s = \Gamma(1-\gamma)H^{\gamma-1} + \sum_{n=0}^{\infty}(-1)^n/n!\zeta(\gamma-n)H^n$ [28]. Therefore we find that, for $P_d(k) \simeq ck^{-\gamma}$,[3]

$$H \simeq \lambda \sum_{n=1}^{\lceil\gamma\rceil-2} \frac{(-1)^{n+1}}{n!}\langle k^n\rangle H^n - \lambda c\Gamma(1-\gamma)H^{\gamma-1} + \mathcal{O}(H^{\lceil\gamma\rceil-1}). \tag{7}$$

If $\gamma > 3$, the H^2 term is the next leading term on the right-hand side of equation (7) and then the critical behavior is given by $H \sim \Delta$ as in the case of all $\langle k^n\rangle$ finite. On the other hand, if $2 < \gamma < 3$, the $H^{\gamma-1}$ term becomes the next leading term and we find that $H \simeq (\lambda/\lambda_c)H - \lambda c\Gamma(1-\gamma)H^{\gamma-1}$. Therefore for $\lambda > \lambda_c$,

$$H \sim \Delta^{1/(\gamma-2)}.$$

[3] There is a logarithmic term $H^{\gamma-1}\ln H$ in equation (7) in case of γ an integer [28].

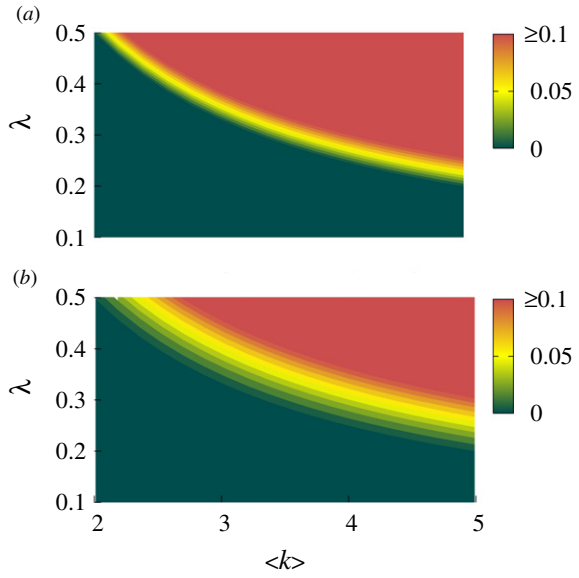

Figure 2. Phase diagram of the Kauffman model for (*a*) a Poisson in-degree distribution and (*b*) a power-law one with the exponent $\gamma = 2.5$, both in the case of uncorrelated in- and out-degree. The color encodes the Hamming distance H obtained by numerically solving equation (3) under setting $H(t) = \bar{H}(t) = H$.

In summary, we can list the values of the critical exponent β varying with the in-degree exponent γ as [18]

$$\beta = \begin{cases} 1 & (\gamma > 3), \\ 1/(\gamma - 2) & (2 < \gamma < 3), \end{cases} \tag{8}$$

which is confirmed numerically (see the inset of figure 1).

The critical exponent β varying with the in-degree distribution as in equation (8) illuminates how the network topology affects the system's response to perturbation. Figure 2 shows phase diagrams of the Kauffman model on model networks with a Poisson in-degree distribution and with a power-law in-degree distribution with the exponent $\gamma = 2.5$, in which color represents the value of the Hamming distance. As shown in the figure, the SF networks with $2 < \gamma < 3$ and thus larger values of β keep the Hamming distance nonzero but small in a much larger region in the $(\lambda, \langle k \rangle)$ plane than those with $\gamma > 3$. Structural and functional organization of cellular networks, parameterized here by $\langle k \rangle$ and λ (p), respectively, may be subject to unexpected changes. Our finding suggests that the systems with strong heterogeneous connectivity patterns can maintain their dynamic criticality robustly, and further, votes for the hypothesis that living organism's machinery lies at the edge of chaos.

4. Boolean networks of finite size

In the thermodynamic limit $N \to \infty$, the critical point does not exhibit any dependence on the network topology. However, for finite N, the critical point itself develops its dependence on

the network topology: it is no more a point but has a nonzero width depending on N. Adopting the finite-size scaling ansatz [29]

$$H = N^{-\beta/\mu}\Psi(\Delta N^{1/\mu}) \tag{9}$$

with the scaling function $\Psi(x) \to$ const for $x \ll 1$ and $\Psi(x) \to x^{\beta}$ for $x \gg 1$, one can see that $H \sim N^{-\beta/\mu}$ in the *critical regime* $|\Delta N^{1/\mu}| \ll 1$. In the λ-axis, this critical regime is $N^{-1/\mu}$ wide for N finite and shrinks to zero in the thermodynamic limit. Therefore, the scaling exponent μ describes the width of the critical regime for finite-size systems. In the critical regime, the cluster of perturbed nodes, explained below, exhibits scale invariance characterized by a power-law distribution of its size, which is connected to the behavior $H \sim N^{-\beta/\mu}$. We show in the next that the asymptotic behavior of the cluster size distribution can be derived using equations (6) and (7), which allows us to obtain the scaling exponent μ and to check equation (9).

4.1. Evolution of perturbed-node clusters

The parameter λ denotes the probability that a node becomes perturbed ($\sigma_i(t+1) \neq \hat{\sigma}_i$) once the configuration of its neighbors is perturbed ($\Sigma_i \neq \hat{\Sigma}_i$). When λ is zero, the Hamming distance becomes zero immediately even though all nodes were perturbed initially. As λ increases from 0, clusters appear, consisting of perturbed nodes that are connected by active edges. An edge from nodes A to B is inactive if the perturbation at node A cannot bring any difference to the dynamical state of node B, and active otherwise. While a node's state can be totally irrelevant to its neighbor connected by an inactive edge, perturbation at one node can propagate to its neighbors through active edges. The perturbed-node clusters evolve with time, decaying or growing. A perturbed node may become normal (non-perturbed) due to its regulators becoming normal at a certain time step, which leads to the decay of the cluster it belonged to. On the other hand, normal nodes may become perturbed due to its regulators becoming perturbed, which leads to the growth of a cluster.

When the parameter λ reaches the critical regime or goes beyond it ($\lambda \gtrsim \lambda_c$), a giant cluster of perturbed nodes appears and contributes to the Hamming distance in the stationary state. There may exist many smaller clusters, but they are hard to survive eventually. A small cluster has a relatively small number of perturbed nodes, and they are surrounded by many normal nodes. Therefore the perturbed nodes in smaller clusters have higher chance of becoming normal than those in larger clusters, which leads to the higher chance of shrinking and decaying of smaller clusters. The perturbed nodes in the giant cluster, on the other hand, have more perturbed nodes as regulators and thus the giant cluster has a higher chance to survive; the probability becomes nonzero when $\lambda \gtrsim \lambda_c$. Therefore the Hamming distance H can be approximated using the size of the largest cluster S via $H \sim S/N$ [30]. We will use this relation in the below to derive the asymptotic behavior of the size distribution of the clusters from the self-consistent equations for H, equations (6) and (7).

4.2. Cluster-size distribution in the annealed approximation

Let us denote the probability that a node belongs to a size-s cluster (of perturbed nodes) by $P(s)$ and consider its generating function $\omega = \mathcal{P}(z) = \sum_s P(s)z^s$. The Hamming distance is related to the generating function as

$$H \simeq \frac{S}{N} \simeq 1 - \sum_{s<S} P(s) \simeq \lim_{N\to\infty}[1 - \mathcal{P}(z_N^*)], \tag{10}$$

where $z_N^* = e^{-1/\tilde{S}}$ and $\tilde{S}$ satisfies $S_2 \ll \tilde{S} \ll S$ with S_2 the second largest cluster size. This relationship, combined with equations (6) and (7), gives us the functional form of the inverse function $z = \mathcal{P}^{-1}(\omega)$. Let us expand the inverse function around $\omega = 1$ as $z = 1 - \sum_{n\geqslant 1} b_n(1-\omega)^n$. Then one can see that this equation should reduce to equations (6) or (7), depending on the in-degree distribution, with z and $1-\omega$ replaced by z_N^* and H, respectively, in the thermodynamic limit. Note that $z_N^* \to 1$ in the thermodynamic limit. Therefore the inverse function $z = \mathcal{P}^{-1}(\omega)$ is expanded around $\omega = 1$ as follows:

$$1 - z \simeq (1 - \lambda/\lambda_c)(1-\omega) + \lambda\langle k^2\rangle(1-\omega)^2/2 + \cdots \tag{11}$$

for $\gamma > 3$ and

$$1 - z \simeq (1 - \lambda/\lambda_c)(1-\omega) + \lambda c\Gamma(1-\gamma)(1-\omega)^{\gamma-1} + \cdots \tag{12}$$

for $2 < \gamma < 3$, where c is the coefficient appearing in the asymptotic behavior of the in-degree distribution $P_d(k) \simeq ck^{-\gamma}$.

The functional behavior of $\mathcal{P}(z)$ around $z = 1$ is then derived from equations (11) and (12). At the critical point ($\lambda = \lambda_c$), it exhibits a singularity depending on the in-degree exponent:

$$1 - \mathcal{P}(z) \sim \begin{cases} (1-z)^{1/2} & (\gamma > 3), \\ (1-z)^{1/(\gamma-1)} & (2 < \gamma < 3). \end{cases} \tag{13}$$

The asymptotic behavior of the cluster-size distribution $P(s)$ can be obtained from the functional form of $\mathcal{P}(z)$ since $P(s) = (1/s!)(d^s/dz^s)\mathcal{P}(z)|_{z=0}$. In particular, when $1 - \mathcal{P}(z) \sim (1-z)^{\tau-1}$ with τ a noninteger, the cluster-size distribution takes a power-law form $P(s) \sim s^{-\tau}$. The asymptotic behavior of $P(s)$ is thus distinguished between $\gamma > 3$ and $2 < \gamma < 3$ as follows:

$$P(s) \sim \begin{cases} s^{-3/2} & (\gamma > 3), \\ s^{-\gamma/(\gamma-1)} & (2 < \gamma < 3). \end{cases} \tag{14}$$

Scale invariance is typical of the system at the critical point and the power-law form of the cluster-size distribution is an example [30]. Our finding shows that the power-law exponent may vary with the degree exponent. In the subcritical ($\lambda < \lambda_c$) or supercritical ($\lambda > \lambda_c$) regime, the linear term in $1-\omega$ in equations (11) and (12) is dominant around $\omega = 1$ and the cluster-size distribution takes an exponential-decaying form like $P(s) \sim e^{-s/s_c}$ [26].

4.3. Scaling exponents

Once the asymptotic behavior of $P(s)$ is known, the scaling property of the largest cluster size may be derived. Using equation (14) in the relation $\sum_{s>S} P(s) \sim S/N$, a part of equation (10), one can see that the size of the largest cluster S scales with the system size N as $S \sim N^{2/3}$ for $\gamma > 3$ and $S \sim N^{(\gamma-1)/\gamma}$ for $2 < \gamma < 3$. It is known that the number of nonfrozen nodes in critical Boolean networks with any fixed number of inputs [31, 32] or fast-decaying in-degree distribution also scales as $N^{2/3}$ [33]. A node i is called nonfrozen if its state (σ_i) is not fixed at either 0 or 1 in the stationary state and thus may be perturbed (but not necessarily). The set of perturbed nodes is thus a subset of the set of nonfrozen nodes and it is noteworthy that their sizes display the same scaling behavior.

The Hamming distance, $H \sim S/N$, in the critical regime is then given by

$$H \sim \begin{cases} N^{-1/3} & (\gamma > 3), \\ N^{-1/\gamma} & (2 < \gamma < 3). \end{cases} \tag{15}$$

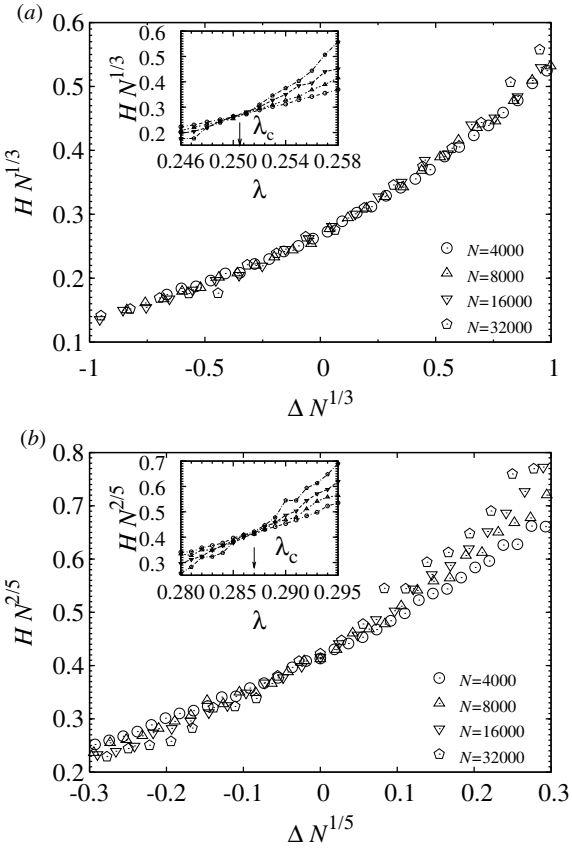

Figure 3. Finite-size scaling behavior in the critical regime. (*a*) Data collapse for different system sizes with $\gamma \to \infty$. The inset shows that $\lambda_c \simeq 0.2505(5)$. (*b*) Data collapse with $\gamma = 2.5$ and $\lambda_c \simeq 0.287(1)$.

Since $H \sim N^{-\beta/\mu}$ in the critical regime from the finite-size scaling ansatz in equation (9), we find that the scaling exponent μ is given by

$$\mu = \begin{cases} 3 & (\gamma > 3), \\ \gamma/(\gamma - 2) & (2 < \gamma < 3). \end{cases} \tag{16}$$

To check numerically the derived finite-size scaling behavior of the Hamming distance, we performed simulations of the Kauffman model on the model networks described in section 3.2 varying the system size. First, we plot the data of $H N^{\beta/\mu}$ as a function of λ for the same average degree ($\langle k \rangle = 4$) and different system sizes ($N = 4000, 8000, 16000, 32000$). They cross at one point, which determines the critical point λ_c according to equation (9). (see the insets of figure 3). Using these values of λ_c, we plotted the same data of $H N^{\beta/\mu}$ versus $\Delta N^{1/\mu}$ in figure 3. The collapse of the data from different system sizes, while a slight deviation is seen in the case of SF networks, presumably due to strong finite-size effects, supports the scaling behavior of the Hamming distance in equation (9) with equations (8) and (16) used.

The finite-size scaling behavior in equation (9) has been identified also in a wide range of dynamical systems on complex networks. In particular, the formation of a giant cluster

in SF networks evolving by adding links has been analyzed through its exact mapping to the $q = 1$ Potts model and found to be characterized by ($\beta = 1, \mu = 3$) for the degree exponent $\gamma > 4$ and ($\beta = 1/(\gamma - 3), \mu = (\gamma - 1)/(\gamma - 3)$) for $3 < \gamma < 4$ [26]. These exponents are very similar to equations (8) and (16). Also in the Ising model [34–37] and the Kuramoto model for synchronization phenomena [38, 39], the exponents are given by ($\beta = 1/2, \mu = 2$) for the degree exponent $\gamma > 5$ and ($\beta = 1/(\gamma - 3), \mu = (\gamma - 1)/(\gamma - 3)$) for $3 < \gamma < 5$. Such similar dependence of the scaling exponents on the degree exponent suggests a common framework to understand the critical phenomena on complex networks [36, 40, 41].

We note that while the scaling plot gives $\lambda_c \simeq 0.251$ for the completely random networks ($\gamma \to \infty$) as predicted by equation (4), the SF networks with $\gamma = 2.5$ have $\lambda_c \simeq 0.289$, deviating from the predicted value 0.25. This deviation seems to be rooted in the use of the annealed approximation for the Hamming distance. It has been reported that the analytic prediction of the critical point in the framework of the mean-field theory deviates slightly from the numerical analysis in the Ising model [37] and in the Kuramoto model for synchronization phenomena [39]. An improvement can be made by considering the Cayley tree with a given degree distribution as the underlying network topology [40].

Our results show that the width of the critical regime $W \sim N^{-1/\mu}$ in the λ-axis increases as the in-degree exponent γ decreases below 3 while its scaling behavior remains the same for all $\gamma > 3$. Since the number of genes in most organisms is not infinite but of order 10^5 at most, the width of the critical regime may be between $\mathcal{O}(10^{-2})$ and $\mathcal{O}(1)$, depending on the in-degree exponent. Such a broad critical regime for small values of γ should help living organisms to remain in the critical regime and in turn, to balance between robustness and evolvability. Individual dynamical responses depend on the properties of the perturbed elements, i.e, on their connectivities and regulating rules, which leads to perturbation propagation on various scales in heterogeneous networks [24]. Here we have analyzed the whole ensemble of such differentiated dynamical responses in heterogeneous networks and found that it can remain critical more easily with the help of extremely heterogeneous connectivity patterns.

The effects of correlation between in- and out-degree of the same node, which we have ignored so far, can be addressed in the results we obtained. In the presence of the in- and out-degree correlation, $(q_k/\langle q \rangle) P_d(k)$ should be considered instead of $P_d(k)$ to compute the Hamming distance, as seen in equation (5). If it holds that $q_k \sim k^{-\theta}$ for large k, we have to consider the effective in-degree exponent, $\gamma_{\text{eff}} = \gamma + \theta$, in place of γ for power-law in-degree distributions in all the results we obtained, including the scaling exponents in equations (8) and (16).

5. Summary and discussion

In summary, we investigated the phase transition between the stable (ordered) and unstable (chaotic) phases in the Boolean dynamical network. Heterogeneous connectivities are found to broaden substantially the small Hamming distance region close to the phase boundary by suppressing the perturbation propagation in the unstable phase. Furthermore, the transition region for finite system sizes turns out to be much wider than in homogeneous networks. Such a robust pseudo-criticality is expected to be also present in transcriptional regulatory networks and also in other biological networks such as neural networks [42], which can be a source for stability and evolvability coexisting in living organisms.

Our results suggest that the heterogeneous connectivity patterns of many biological networks have been selected in the course of evolution in part to serve for achieving stability and evolvability simultaneously. Therefore, it would be desirable to propose a model in which heterogeneous connectivity patterns emerge driven by the evolutionary pressure towards a

broad edge of chaos. There are indeed models for co-evolution of network structure and dynamics, which reproduce networks with a selected number of links supporting dynamic criticality by dynamics-correlated addition and deletion of links [43, 44]. Similarly to these models, if one allows link rewiring only, preserving the total number of links, and make it happen depending on the network's dynamical state, the network is expected to be organized so as to have a broad degree distribution, which is under investigation.

While we focussed on the Hamming distance to capture the effects of structural features on the dynamic stability of Boolean networks, it would be also interesting to see how the heterogeneous connectivity patterns affect the properties of attractors in the configuration space, given the recent interest [10–14] in the scaling behavior of the number and length of the attractors in the critical Boolean networks. It has been shown that the median attractor length of SF networks is larger than that of networks with fixed in-degree at the critical point [45, 46], but much more properties remain to be investigated.

The asymptotic behaviors of the in-degree distributions of real transcriptional regulatory networks of *E. coli* [5] and yeast [2, 21] are hard to discern due to finite-size effects. In both organisms, there are about one hundred regulators (nodes with outgoing links), which impose a cut-off in the measured in-degree distributions. However, one can find for the network of yeast that its in-degree distribution is much broader than that of the networks generated by randomly rewiring the links, and that the Hamming distance of the Boolean dynamics is much slower than that in the randomized networks, demonstrating the contribution of heterogeneous connectivity pattern to maintaining dynamic criticality [18].

It should be noted that the real transcriptional regulatory networks and other biological networks have much richer structural properties than described here and their relation to the dynamic criticality of the system is of interest. For example, the correlation of the degrees of neighboring nodes has been identified in many real-world networks [47, 48] including the yeast gene regulatory network [20] and there are studies on the effects of the degree–degree correlation on the structure and dynamics of complex networks [49–51]. While it has been shown [51] that a negative (positive) degree–degree correlation is irrelevant (relevant) to the percolation transition in complex networks, it still remains to be addressed how the correlation affects the critical phenomena of Boolean networks.

References

[1] Thieffry D, Huerta A M, Pérez-Rueda E and Collado-Vides J 1998 *Bioessays* **20** 433
[2] Lee T I *et al* 2002 *Science* **298** 799
[3] Dobrin R, Beg Q K, Barabási A-L and Oltvai Z N 2004 *BMC Bioinformatics* **5** 10
[4] Guelzim N, Bottani S, Bourgine P and Képès F 2002 *Nature Genetics* **31** 60
[5] Shen-Orr S S, Milo R, Mangan S and Alon U 2002 *Nature Genetics* **31** 64
[6] Babu M M *et al* 2004 *Curr. Opin. Struct. Biol.* **14** 283
[7] Kauffman S 1969 *J. Theor. Biol.* **22** 437
Kauffman S 1993 *The Origins of Order: Self-organization and Selection in Evolution* (Oxford: Oxford University Press)
[8] Derrida B and Pomeau Y 1986 *Europhys. Lett.* **1** 45
[9] Aldana M, Coppersmith S and Kadanoff L P 2003 Perspectives and problems in nonlinear science: a celebratory volule in honor of Lawrence Sirovich *Springer Applied Mathematical Sciences Series* ed E Kaplan, J E Marsden and K R Sreenivasan
[10] Bastolla U and Parisi G 1998 *Physica* D **115** 203
Bastolla U and Parisi G 1998 *Physica* D **115** 219
[11] Bilke S and Sjunnesson F 2001 *Phys. Rev.* E **65** 016129
[12] Samuelsson B and Troein C 2003 *Phys. Rev. Lett.* **90** 098701
[13] Drossel B, Mihaljev T and Greil F 2005 *Phys. Rev. Lett.* **94** 088701
[14] Klemm K and Bornholdt S 2005 *Phys. Rev.* E **72** 055101

[15] Harris S E, Sawhill B K, Wuensche A and Kauffman S 2002 *Complexity* **7** 23
[16] Kauffman S, Peterson C, Samuelsson B and Troein C 2003 *Proc. Natl Acad. Sci. USA* **100** 14796
[17] Albert R and Barabási A-L 2002 *Rev. Mod. Phys.* **74** 47
[18] Lee D-S and Rieger H 2007 *J. Theor. Biol.* **248** 618
[19] Kauffman S, Peterson C, Samuelsson B and Troein C 2004 *Proc. Natl Acad. Sci. USA* **101** 17102
[20] Balcan D, Kabakçioğlu A, Mungan M and Erzan A 2005 *PLoS One* **2** e501
[21] Luscombe N M, Babu M M, Yu H, Snyder M, Teichmann S A and Gerstein M 2004 *Nature* **431** 308
[22] Moreira A A and Amaral L A N 2005 *Phys. Rev. Lett.* **94** 218702
[23] Oosawa C and Savageau M A 2002 *Phyisca* D **170** 143
[24] Aldana M and Cluzel P 2003 *Proc. Natl Acad. Sci. USA* **100** 8710
[25] Lee D-S 2008 unpublished data
[26] Lee D-S, Goh K-I, Kahng B and Kim D 2004 *Nucl. Phys.* B **696** 351
[27] Gumbel E J 1958 *Statistics of Extremes* (New York: Columbia University Press)
[28] Robinson J E 1951 *Phys. Rev.* **83** 678
[29] Marro J and Dickman R 1999 *Nonequilibrium Phase Transitions in Lattice Models* (Cambridge: Cambridge University Press)
[30] Stauffer D and Aharony A 1994 *Introduction to Percolation Theory* (London: Taylor and Francis)
[31] Kaufman V, Mihaljev T and Drossel B 2005 *Phys. Rev.* E **72** 046124
[32] Mihaljev T and Drossel B 2006 *Phys. Rev.* E **74** 046101
[33] Samuelsson B and Socolar J E 2006 *Phys. Rev.* E **74** 036113
[34] Aleksiejuk A, Holyst J A and Stauffer D 2002 *Physica* A **310** 260
[35] Leone M, Vázquez A, Vespignanai A and Zecchina R 2002 *Eur. Phys. J.* B **28** 191
[36] Iglói F and Turban L 2002 *Phys. Rev.* E **66** 036140
[37] Herrero C P 2004 *Phys. Rev.* E **69** 067109
[38] Hong H, Choi M Y and Kim B J 2002 *Phys. Rev.* E **65** 026139
[39] Lee D-S 2005 *Phys. Rev.* E **72** 026208
[40] Dorogovtsev S N, Goltsev A V and Mendes J F F 2002 *Phys. Rev.* E **66** 016104
[41] Hong H, Ha M and Park H 2007 *Phys. Rev. Lett.* **98** 258701
[42] Bornholdt S and Röhl T 2003 *Phys. Rev.* E **67** 066118
[43] Bornholdt S and Rohlf T 2000 *Phys. Rev. Lett.* **84** 6114
[44] Liu M and Bassler K E 2006 *Phys. Rev.* E **74** 041910
[45] Iguchi K, Kinoshita S and Yamamda H S 2007 *J. Theor. Biol.* **247** 138
[46] Kinoshita S, Iguchi K and Yamamda H S 2008 *AIP Conf. Proc.* **982** 2128
[47] Pastor-Satorras R, Vazquez A and Vespignani A 2001 *Phys. Rev. Lett.* **87** 258701
[48] Newman M E J 2002 *Phys. Rev. Lett.* **89** 208701
[49] Vazquez A and Moreno Y 2003 *Phys. Rev.* E **67** 015101
[50] Bianconi G and Marsili M 2007 *Phys. Rev.* E **76** 026116
[51] Noh J D 2007 *Phys. Rev.* E **76** 026116